

Clase 6: Modelos de series de tiempo.

Modelos de toma de decisiones

Paul Castillo Bardález

Maestría en Finanzas

Marzo 2012

Motivación modelos de series de tiempo

¿Cuándo debemos utilizar modelos de series de tiempo?

- Cuando se requiere realizar pronósticos de corto plazo. Los modelos de series de tiempo utilizan información histórica de manera eficiente por lo que permiten realizar pronósticos rápidos y a bajo costo.
- Cuando se requiere medir la relación dinámica entre dos o más variables. Por ejemplo, la relación entre volatilidad y aversión al riesgo, o entre inflación y tasa de interés de política monetaria.
- Cuando se requiere cuantificar la importancia de choques en la dinámica de una o más variables

Motivación modelos de series de tiempo

Tipos de modelos

- En el curso vamos a estudiar varios tipos de modelos de series de tiempo:
 - Univariados lineales: utilizan información histórica de una sola serie de tiempo, **modelos ARIMA**.
 - Multivariados lineales: Utilizan información de varias series de tiempo. Tienen la ventaja de utilizar la interacción dinámica entre varias variables, **modelos VAR**.
 - Univariados no lineales, permiten modelar la dinámica de la varianza de retornos, **modelos GARCH**
 - Modelos de largo plazo: se utilizan para medir la relación de largo plazo entre variables, **modelos cointegrados, y de corrección de errores**

Modelos ARIMA

Definiciones Básicas en series de tiempo

- **Proceso estocástico:** Se define como una secuencia o conjunto de variables aleatorias ordenadas en el tiempo.
$$\{Y_t\}_{t=-\infty}^{t=\infty} = \{\dots, Y_{-1}, Y_0, Y_1, \dots\}$$
- **Serie de tiempo:** Es una realización particular de un proceso estocástico. Ejemplo, la tasa de crecimiento del producto anual para Perú entre los años 1992 y 2011.
- En lo que sigue vamos a estudiar como modelar el comportamiento de las series de tiempo utilizando una variedad de herramientas.

Modelos ARIMA

Definición de estacionariedad

- Dos tipos a) Débil y fuerte
- **Débil:** Un proceso estocástico, $\{y_t\}$ se dice que es estacionario en el sentido débil si sus primeros momentos centrales son independientes del tiempo.

$$\begin{aligned} E(y_t) &= \mu & E((y_t - \mu)^2) &= \sigma^2 \\ E((y_t - \mu)(y_{t-j} - \mu)) &= \gamma_j \end{aligned}$$

Donde μ , σ^2 y γ_j representan la media, varianza y covarianza de orden j del proceso estocástico y_t

- **Fuerte:** Un proceso estocástico es estacionario en el sentido fuerte si las correspondientes distribuciones de probabilidad de cada y_t son independientes del tiempo, dependen únicamente de la distancia s entre y_t y y_{t-s}

- **Ruido Blanco:** Es el proceso estocástico más elemental. $y_t = \varepsilon_t$, donde, $E(y_t) = 0$

$$E\left((y_t - \mu)^2\right) = \sigma^2 \quad E\left((y_t - \mu)(y_{t-j} - \mu)\right) = 0$$

- **Paseo aleatorio:** $y_t = y_{t-1} + \varepsilon_t$, donde ε_t es un ruido blanco. Este proceso no es estacionario en varianzas. Mediante substitución

repetitiva se obtiene que $y_t = y_{-1} + \sum_{s=1}^{s=t} \varepsilon_s$, $E(y_t) = 0$,

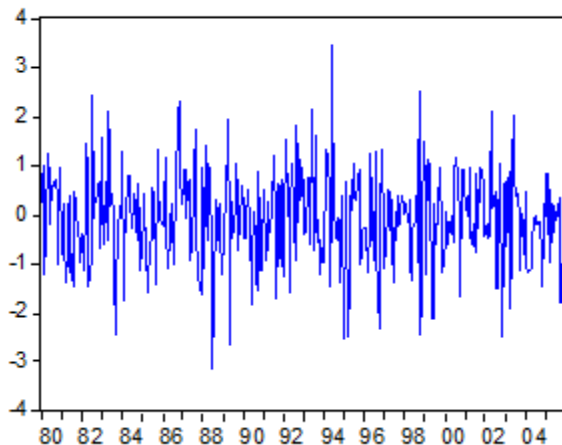
$E(y_t^2) = t\sigma^2$, donde y_{-1} esta fijo. Aplicando el operador diferencias a y_t , se transforma el proceso en estacionario, $\Delta y_t = \varepsilon_t$

- **Tendencia determinística:** $y_t = \alpha + \beta t + \varepsilon_t$, donde ε_t es un ruido blanco. Este proceso no es estacionario en medias, $E(y_t) = \alpha + \beta t$, para transformar y_t en un proceso estacionario, hay que quitarle la media, $X_t = y_t - \alpha - \beta t = \varepsilon_t$

Modelos ARIMA

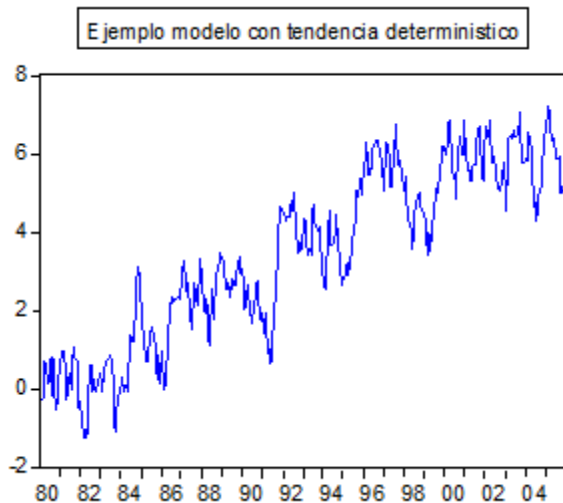
Modelo ruido blanco

Ejemplo de un Ruido Blanco Gaussiano



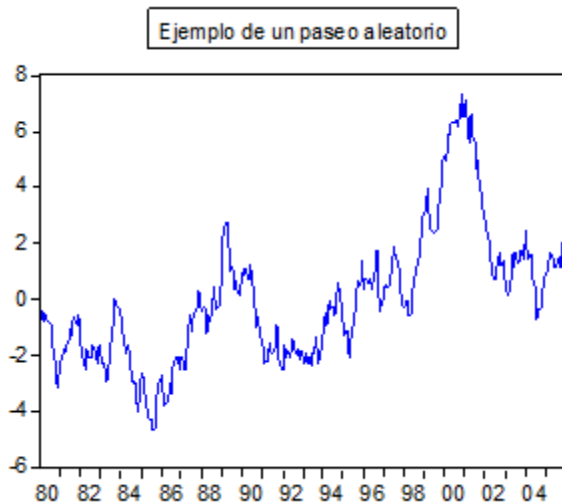
Modelos ARIMA

Tendencia determinística



Modelos ARIMA

Paseo aleatorio



- Se identifica con la letra L , es un operador matemático lineal que se aplica sobre series de tiempo, posee las siguientes propiedades. Sea Y_t una serie de tiempo, entonces se cumple que,
 - $LY_t = Y_{t-1}$
 - $L^n Y_t = Y_{t-n}$
 - $a + bLY_t = a + bY_{t-1}$
 - $L^{-1}Y_t = Y_{t+1}$

- **Condiciones de estacionariedad:** Un proceso estocástico AR(p) es estacionario si todas las raíces del polinomio de rezagos, $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p$ están fuera del círculo unitario. $\Phi(L) = (1 - \lambda_1 L)(1 - \lambda_2 L)(1 - \lambda_3 L) \dots (1 - \lambda_p L)$, raíces fuera del círculo unitario, $L = \left| \frac{1}{\lambda_i} \right| > 1$, implican $|\lambda_i| < 1$

Modelos ARIMA

Representación de medias Móviles de un modelo

$AR(p)$

- A partir de la representación de medias móviles del modelo $AR(p)$ se puede obtener:
 - El multiplicador dinámico, $\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s$
 - La función impulso respuesta, es el gráfico, ψ_s versus s
 - El impacto acumulado del choque en el largo plazo, $\sum \psi_s = \Psi(1)$
 - Los segundos momentos de y

$$E \left((y_t - \mu)^2 \right) = \sigma^2 \sum_{h=0}^{h=\infty} \psi_h^2$$

$$E \left((y_t - \mu) (y_{t-s} - \mu) \right) = \sigma^2 \sum_{h=0}^{h=\infty} \theta_h \theta_{h+s}$$

Modelos ARIMA

Ejemplo: Modelo AR(1)

- Este modelo se representa por

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

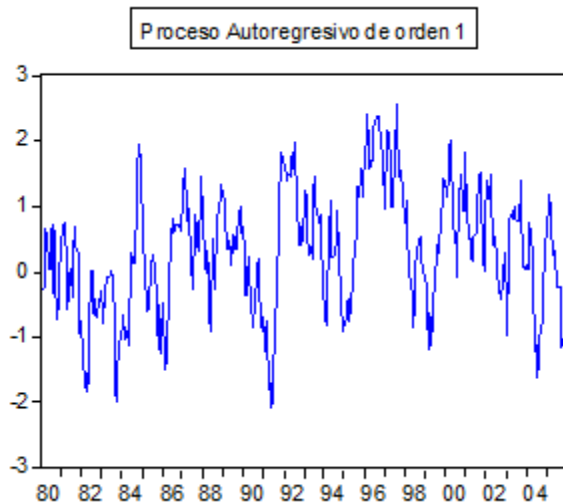
- Este proceso es estacionario si $|\phi_1| < 1$. En este caso es posible expresar el modelo autorregresivo como un proceso de medias móviles infinito.

$$y_t = \sum_{h=0}^{h=\infty} \psi_h \varepsilon_{t-h}$$

- En este caso, $\psi_k = \phi_1^k$.

Modelos ARIMA

Ejemplo: Modelo AR(1)



- Este modelo esta representado por:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \dots \theta_q \varepsilon_{t-q}$$

- Sus correspondientes primeros y segundos momentos estan dados por

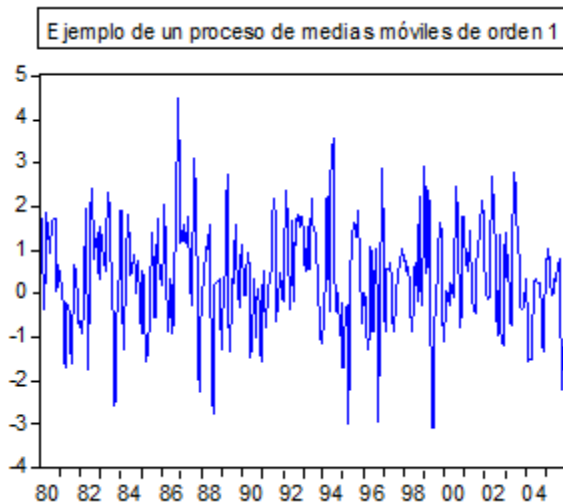
$$E(y_t) = \mu, \quad E((y_t - \mu)^2) = \sigma^2 \sum_{k=0}^{k=q} \theta_k^2$$

$$E((y_t - \mu)(y_{t-s} - \mu)) = \sigma^2 \sum_{k=0}^{k=q} \theta_k \theta_{k+s}$$

donde, $E((y_t - \mu)(y_{t-s} - \mu)) = 0$, para $s > q$.

Modelos ARIMA

Ejemplo: Modelo MA(1)



Modelos ARIMA

La representación completa

- $ARMA(p, q)$. Un proceso $ARMA(p, q)$ incluye tanto una parte autorregresiva como una de medias móviles. $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} \dots \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \dots \theta_q \varepsilon_{t-q}$
- Expresando el proceso en términos de operadores de rezagos tenemos que: $\Phi(L)y_t = c + \Theta(L)\varepsilon_t$
- $ARIMA(p, d, q)$ es un proceso $ARMA(p, q)$ para la diferencia de orden d de y_t , esto es, $\Phi(L)\Delta^d y_t = c + \Theta(L)\varepsilon_t$, donde $\Delta^d = (1 - L)^d$

Modelos ARIMA

Definición

- Los modelos $ARIMA(p,d,q)$ son modelos $ARMA(p,q)$ para la diferencia de orden d de la serie de tiempo
- Sea Y_t una serie de tiempo no estacionaria, entonces, ΔY_t se define como el operador diferencia, tal que, $\Delta = (1 - L)$, entonces, $\Delta Y_t = Y_t - Y_{t-1}$. También, $\Delta^2 Y_t = \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Ejemplos, $ARIMA(1,1,1)$ esta determinado por, $(1 - \phi_1 L) \Delta Y_t = (1 + \theta_1 L) \varepsilon_t$, que en su forma extensiva puede escribirse como, $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$,

$$Y_t = (1 + \phi_1) Y_{t-1} - \phi_1 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1},$$

Modelos ARIMA

Modelos que incluyen estacionalidad

- La estacionalidad se define como fluctuaciones recurrentes en las series de tiempo asociadas a eventos periodicos que se repiten cada año, como por ejemplo, gastos por navidad, semana santa, o ventas en verano o invierno. Estos eventos generan correlaciones especiales, con los datos de la misma fecha en el año previo,
- Ejemplos, $SAR(1) \ (1 - \phi_1^s L^4) Y_t = \varepsilon_t$, $SMA(1) \ Y_t = (1 + \theta_1^s L^4) \varepsilon_t$
- Considerando estacionalidad, se pueden definir modelos más complejos, por ejemplo, $ARIMA(p, p_s, d, d_s, q, q_s)$, donde, los parámetros con subíndice s representan correlaciones estacionales, por ejemplo, $ARIMA(1, 1, 1, 1, 1, 1)$, para datos trimestrales estaría dado por

$$(1 - \phi_1^s L^4) (1 - \phi_1 L) \Delta \Delta^s Y_t = (1 + \theta_1^s L^4) (1 + \theta_1 L) \varepsilon_t$$

donde, $\Delta^s = (1 - L^4)$

Proposition

- i) *(descomposición de Wold). Cualquier proceso estacionario en covarianzas, y_t puede representarse de la forma:*

$$y_t = c + \sum_{k=0}^{\infty} \psi_k \omega_{t-k}$$

donde: $\psi_0 = 1$ y $\sum_{k=0}^{\infty} \psi_k^2 < \infty$. El término ω_t es un ruido blanco y representa el error de la predicción lineal de y_t sobre sus valores rezagados $\omega_t = y_t - E_t(y_t / y_{t-1}, y_{t-2} \dots y_{t-p})$

Modelos ARIMA

Identificación de modelos ARIMA(p,d,q)

- Método iterativo de estimación y chequeo de las propiedades del modelo. Este proceso puede resumirse en los siguientes cuatro pasos:
 - Transformar los datos buscando inducir estacionariedad. Si la tendencia es determinística, se remueve estimando para la serie y_t un polinomio determinístico de la forma:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 \dots \alpha_s t^s + \varepsilon_t$$

el residuo de esta regresión se considera como el proceso estacionario que se esta buscando identificar.

- Si la tendencia no es determinística se procede a diferenciar la serie tantas veces sea necesario para convertirla en estacionaria.

$$\Phi(L)\Delta^d y_t = \Theta(L)$$

- Para determinar si se ha alcanzado estacionariedad se debe examinar las FAS y FAP de la serie transformada. Si esta es estacionaria, ambas debería aproximarse a cero rápidamente.

Modelos ARIMA

Identificación de modelos ARIMA(p,d,q)

- Utilizar p_{\max} y q_{\max} del modelo ARMA(p,q) que mejor puede describir la serie.
- Estimar los parámetros de los polinomios de rezagos $\Phi(L)$ y $\Theta(L)$
- Hacer un diagnóstico del residuo Si el modelo asumido es el correcto, entonces
 - El residuo no debería presentar correlación serial,
 - Debe ser homocedástico, e idealmente debería tener media cero y distribuirse normalmente.

Modelos ARIMA

Identificación de modelos ARIMA(p,d,q)

- De no cumplirse algunas de estas propiedades del residuo, se regresa al paso (2) y se repite el proceso. Adicionalmente, se puede utilizar criterios de información.
- El criterio de información de Akaike (AIC)

$$AIC = -2\frac{L}{T} + 2\frac{k}{T}$$

El criterio de Schwarz (SC)

$$SC = -2\frac{L}{T} + k\frac{\log(T)}{T}$$

Y el criterio de Hanna-Quinn (HQ)

$$HQ = -2\frac{L}{T} + 2k\frac{\log(\log(T))}{T}$$

Donde, T : tamaño de muestra, L el logartimo de la función de verosimilitud, mayor este valor, mayor la probabilidad de que los datos prevengan del modelo planteado, y finalmente, k que representa el número de parámetros estimados

Modelos ARIMA

Funciones de Autocorrelación Simple (FAS) y Parcial(FAP)

- La FAS se define como $\rho_s = \frac{\gamma_s}{\gamma_0}$, El correlograma se define como el gráfico de $\hat{\rho}_s$ versus s
- El estadístico de Box-Pierce y Box-Ljung Q: Permiten probar la hipótesis conjunta: $\rho_1 = \rho_2 = \dots = \rho_s = 0$, el estadístico de BP está dado por , $Q(k) = T \sum_{s=1}^k \hat{\rho}_s^2$, donde, $Q(k) \sim \chi^2(k)$
- Box-Ljung ajusta el estadístico utilizando grados de libertad para mejorar el desempeño del estadístico en muestras pequeñas, $Q(k) = T(T+2) \sum_{s=1}^k \frac{\hat{\rho}_s^2}{T-s}$, donde, $Q(k) \sim \chi^2(k)$

Modelos ARIMA

La función de Autocorrelación Parcial(FAP)

- Es la correlación entre y_t y y_{t-s} es la correlación simple entre y_{t-s} y y_t menos la parte linealmente explicada por los rezagos intermedios.
$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} \dots \phi_p y_{t-p}$$
- La función de autocorrelación parcial esta dada por el gráfico de ϕ_k versus k
- Para un AR(p) $\phi_k = 0$ para $k > p$
- Para MA(q) $\rho_k = 0$ para $k > q$

- La predicción de y_t , h periodos en el futuro condicional en la información I_t se define como

$$y_{t+1/t} = \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} \dots \phi_p y_{t-p+1}$$

- Su correspondiente error de predicción es : $\epsilon_{t+h/t} = y_{t+h} - y_{t+h/t}$
- El error cuadrático medio del error de predicción esta dado por:

$$MSE(\epsilon_{t+h/t}) = E(\epsilon_{t+h/t}^2) = E[(y_{t+h} - y_{t+h/t})^2]$$

- El correspondiente error de predicción esta dado por:

$$y_{t+1/t} = \phi_1 y_t$$

$$\hat{\epsilon}_{t+1/t} = y_{t+1} - y_{t+1/t} = \epsilon_{t+1}$$

$$\hat{\epsilon}_{t+2/t} = y_{t+2} - y_{t+2/t} = \epsilon_{t+2} + \psi \epsilon_{t+1}$$

$$\hat{\epsilon}_{t+h/t} = y_{t+h} - y_{t+h/t} = \epsilon_{t+h} + \psi \epsilon_{t+h-1} \dots \psi_{t+1}^{h-1} \epsilon_t$$

Modelos ARIMA

Evaluando la predicción

- Dado un horizonte de predicción, h , y n valores predichos
- La raíz del error cuadrático medio se define como:

$$RECM = \sqrt{\sum_{t=T+1}^{T+n} \frac{(\hat{y}_t - y_t)^2}{n}}$$

- El error absoluto medio

$$MAE = \sum_{t=T+1}^{T+n} \frac{|(\hat{y}_t - y_t)|}{n}$$

- La media absoluta del porcentaje de error

$$MAPE = 100 \sum_{t=T+1}^{T+n} \frac{|(\hat{y}_t - y_t) / y_t|}{n}$$

- El coeficiente de desigualdad de Theil

$$UTHEIL = \frac{\sqrt{\sum_{t=T+1}^{T+n} \frac{(\hat{y}_t - y_t)^2}{n}}}{\sqrt{\sum_{t=T+1}^{T+n} \frac{(\hat{y}_t)^2}{n}} \sqrt{\sum_{t=T+1}^{T+n} \frac{(y_t)^2}{n}}}$$

- Los dos primeros estadísticos dependen de la escala de la variable dependiente. Estos deberían usarse para comparar predicciones de la misma serie para diferentes modelos.
- Los dos últimos indicadores son independientes de la escala. La UTHEIL siempre entre cero y uno, cuando es cero indica un ajuste perfecto.

- ¿Cuál es la respuesta en 4 trimestres de la inflación a un incremento de 100 puntos en la tasa de interés de política monetaria?.
- ¿Cuál es la correlación entre el tipo de cambio real y el PBI en un horizonte de 4 trimestres si los choques que afectan la economía son de oferta?,
- ¿Qué choques son los más importantes en la varianza de la inflación en el corto plazo y en el largo plazo?
- ¿Cómo hubiera respondido el PBI si la tasa de interés internacional no se hubiera incrementado?

Modelos VAR

Un ejemplo simple

- Sea $Y_t' = \begin{bmatrix} x_t & \pi_t & i_t \end{bmatrix}$ un vector de tres variables endógenas, brecha producto, inflación y la tasa de interés nominal. Un modelo VAR(p), tiene la siguiente forma,

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + A_3 Y_{t-3} + \dots A_p Y_{t-p} + \varepsilon_t$$

$$\text{donde, } A_1 = \begin{bmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix} \text{ y}$$

$$\Sigma = E_t(\varepsilon_t \varepsilon_t') = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{31} & \Sigma_{33} \end{bmatrix}$$

- Los datos determinan las propiedades de Σ y p . La teoría económica determina la estructura económica consistente con Σ , en particular, $\varepsilon_t = B_0^{-1} V_t$, B_0 no esta determinado para los datos.

- Sea $Y_t' = [y_{1,t} \dots y_{k,t}]$ un vector de variables endógenas, el modelo VARE esta representado por,

$$B_0 Y_t = B_1 Y_t + B_2 Y_{t-2} \dots + B_p Y_{t-p} + V_t$$

$$E_t (V_t V_t') = \Omega$$

- Alternativamente, utilizando el operador de rezagos el modelo VAR se puede representar como:

$$B(L) Y_t = V_t$$

$$B(L) = B_0 - B_1 L - B_2 L^2 \dots - B_p L^p$$

- Además Ω es un matriz diagonal de tamaño $k \times k$ y B_s son matrices también de orden $k \times k$ con elemento típico, b_{ij}^s

Modelos VAR

¿Por qué estructural?

- Los choques son ortogonales $E v_{i,t} v_{j,t} = 0$
- Los choques tienen interpretación económica (choques de demanda, política monetaria, gasto público, etc)
- $B_0 \neq I$ mide relaciones contemporáneas entre variables

- Tres variables, $Y'_t = [\pi_t \quad i_t \quad \mu_t]$
- Ecuación estructural, la regla de Taylor, $i_t = 1.5\pi_t - 1.25\mu_t$
- π_t : inflación, i_t : tasa de interés nominal, y μ_t : desempleo,
- En términos matriciales

$$Y'_t = [\pi_t \quad i_t \quad \mu_t]$$

$$V'_t = [\eta_t \quad v_t \quad g_t]$$

$$B_0 = \begin{bmatrix} 1 & * & * \\ -1.5 & 1 & 1.25 \\ * & * & 1 \end{bmatrix}$$

Modelos VAR

Un modelo simple de inflación

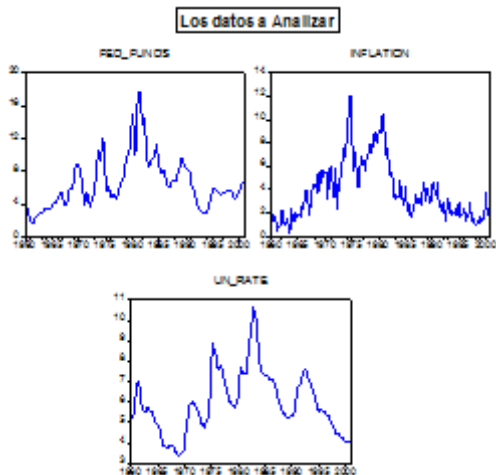
- La forma reducida se obtiene multiplicando el sistema por B_0^{-1}

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} \dots + A_p Y_{t-p} + \tilde{\zeta}_t$$

$$E_t (\tilde{\zeta}_t \tilde{\zeta}_t') = \Sigma$$

- Donde, $\Sigma = B_0^{-1} \Omega B_0^{-1'}$ y $A_s = B_0^{-1} A_s$
- En este caso, $E_t \tilde{\zeta}_{i,t} \tilde{\zeta}_{jt} \neq 0$ los choques están correlacionados

- Tres variables, $Y'_t = \begin{bmatrix} \pi_t & i_t & \mu_t \end{bmatrix}$



- El polinomio de rezagos, $Y_t = \zeta_t + \Psi_1 \zeta_{t-1} + \Psi_2 \zeta_{t-2} + \Psi_3 \zeta_{t-3} \dots$, es la representación de medias móviles de forma reducida
- La representación de medias móviles de forma estructural se obtiene reemplazando, $\zeta_t = B_0^{-1} V_t$, de tal forma que $\Phi(L) = \Psi(L) B_0^{-1}$, de donde se obtiene,

$$Y_t = \Phi_0 V_t + \Phi_1 V_{t-1} + \Phi_2 V_{t-2} + \Phi_3 V_{t-3} \dots$$

- Las funciones impulso respuesta, la descomposición de varianza, los momentos condicionados y la descomposición histórica se determina a partir de la representación de medias móviles estructural.

- Mide la respuesta de las variables endógenas s periodos a futuro al aumento por una sólo vez del choque j , esto es $\frac{\partial Y_{t+s}}{\partial v_{j,t}}$
- De la representación de medias móviles, $\frac{\partial Y_{t+s}}{\partial v_{j,t}} = \Phi_s(:, j)$
- Si el VAR es estacionario en covarianzas, $\Psi_s(:, j) \rightarrow 0$, el impacto de largo plazo de los choques esta determinado por $\Phi(1) = I + \Phi_1 + \Phi_2 + \Phi_3 \dots$

- Se obtiene de la representación de medias móviles estructurales,

$$Y_t = \Phi(L) V_t,$$
- Particionando apropiadamente el polinomio de rezagos $\Phi(L)$,
 tenemos que,

$$Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} \Phi^{11}(L) & \Phi^{12}(L) & \dots & \Phi^{1k}(L) \\ \Phi^{21}(L) & \Phi^{22}(L) & \dots & \Phi^{2k}(L) \\ \vdots & \vdots & \vdots & \vdots \\ \Phi^{k1}(L) & \Phi^{k2}(L) & \dots & \Phi^{kk}(L) \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{k,t} \end{bmatrix}$$

- Entonces, los segundos momentos condicionados se definen como: $E_t(y_{i,t}/v_{j,t}) = \Phi^{ij}(1)^2 \sigma_j^2$, la covarianza condicional $E_t(y_{i,t}y_{s,t}/v_{j,t}) = \Phi^{ij}(1)\Phi^{is}(1)\sigma_j\sigma_s$

- ¿Que tan importante es la varianza del choque estructural j en el error de predicción s periodos a futuro?
- Esto esta determinado por

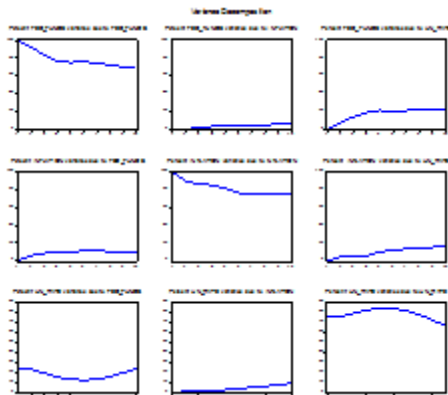
$$f_{t+s} = \Sigma + \Psi_1 \Sigma \Psi_1' + .. \Psi_{s-1} \Sigma \Psi_{s-1}'$$

- Sabemos que $\xi_t = B_0^{-1} V_t$, alternativamente,

$$\xi_t = \begin{bmatrix} b_0^1 & b_0^2 & \dots & b_0^k \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{k,t} \end{bmatrix}, \quad \xi_t = b_0^1 v_{1,t} + b_0^2 v_{2,t} .. b_0^k v_{k,t}$$

- Utilizando esta notación, $\Sigma = b_0^1 b_0^{1'} \sigma_1^2 + b_0^2 b_0^{2'} \sigma_2^2 .. b_0^k b_0^{k'} \sigma_k^2$

- En el caso del ejemplo,



Pruebas de raíz unitaria

Introducción

- Las pruebas de raíces unitarias son pruebas formales que evalúan la estacionariedad de procesos estocásticos que admiten una representación autorregresiva.
- La existencia de raíces unitarias implica que la serie de tiempo contiene una tendencia estocástica, y por lo tanto, los choques a la serie tienen efectos permanentes.
- Regresiones de MCO entre series que no son estacionarias y que no cointegran pueden ser espúreas

Pruebas de raíz unitaria

La prueba de Dickey-Fuller

- Una prueba de raíz unitaria consiste en probar si algunas de las raíces de $\alpha(L)$ es unitaria en $\alpha(L)$ $y_t = \varepsilon_t$, donde,
 $\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 \dots \alpha_p L^p$
- En el caso del modelo AR(1), las hipótesis a probar serían

$$H_0: \alpha = 1, \text{ que equivale a } y_t \sim I(1)$$

$$H_a: \alpha < 1, \text{ que equivale a } y_t \sim I(0)$$

- El estadístico en este caso es, $t_{\alpha=1} = \frac{\hat{\alpha}-1}{se(\hat{\alpha})}$, donde, $\hat{\alpha}$ es el estimador de MCO de α
- En el caso de que la hipótesis nula sea cierta, la distribución de $\hat{\alpha}$, no es una normal sino una distribución tipo Dickey-Fuller, los momentos de esta distribución no convergen a constantes, sino a funciones aleatorias denominadas procesos Brownianos. Estos procesos son la versión límite de un paseo aleatorio.

Pruebas de raíz unitaria

La prueba de Dickey-Fuller

- Conceptualmente las pruebas de raíces unitarias son sencillas, equivalen a una prueba de significancia tipo t para $\hat{\alpha}$. Sin embargo, existen varios problemas.
- Bajo la hipótesis nula la distribución de $t_{\alpha=1} = \frac{\hat{\alpha}-1}{se(\hat{\alpha})}$ es no estándar. No tiene distribuciones de probabilidad que admiten representación analítica, por lo que los valores críticos se deben calcular mediante métodos de simulación.
- Estas distribuciones dependen de la inclusión de elementos determinísticos, constantes, tendencias y variables dummies.

Pruebas de raíz unitaria

La prueba de Dickey-Fuller

- Phillips demostró que :
 - $\hat{\alpha}$ es superconsistente en el sentido que $T(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, (1 - \alpha^2))$, esto es $\hat{\alpha}$ converge a su valor poblacional a una tasa T mayor de lo usual, $T^{\frac{1}{2}}$
 - La distribución límite se denomina distribución Dickey-Fuller.

El tratamiento de los componentes determinísticos

- Cuando se realizan pruebas de raíces unitarias es crucial especificar adecuadamente el componente determinístico
- Las propiedades determinísticas de los datos bajo la hipótesis alternativa determina el tipo de prueba a emplearse.
- El tipo de componente determinístico determina las propiedades asintóticas de la prueba de raíz unitaria.

Pruebas de raíz unitaria

La prueba de Dickey-Fuller

- Esta prueba es válida si el verdadero proceso es un modelo AR(1)
- Los valores críticos de la prueba dependen de los componentes determinísticos. Se consideran tres casos:
 - Caso I: sin constante ni tendencia, $y_t = \alpha y_{t-1} + \varepsilon_t$
 - Caso II; con constane, $y_t = c + \alpha y_{t-1} + \varepsilon_t$
 - Caso II: con constante y tendencia, $y_t = c + \delta t + \alpha y_{t-1} + \varepsilon_t$

Prueba de Raíz unitaria de Dickey-fuller

Caso II: Constante solamente

- En este caso,

$$H_0: \alpha = 1, c = 0, y_t \sim I(1)$$

$$H_a: \alpha < 1, y_t \sim I(0)$$

- Esta formulación es apropiada para series que no muestran tendencia como tasa de interes, variaciones en el tipo de cambio y spreads.
- Los valores críticos en este caso se desplazan hacia la izquierda de la distribución, así el valor crítico al 5 por ciento de significancia es -2.861 y no -1.645 como cuando se rechaza la hipótesis nula.

Prueba de Raíz unitaria de Dickey-fuller

Caso II: Constante y tendencia

- En este caso,

$$H_0: \alpha = 1, \delta = 0, y_t \sim I(1)$$
$$H_a: \alpha < 1, y_t \sim I(0) \text{ con tendencia determinística}$$

- Esta formulación es apropiada para series que muestran tendencia como precios de activos financieros o agregados macroeconómicos, como el PBI.
- En este caso también los valores críticos se desplazan hacia la izquierda de la distribución, así el valor crítico al 5 por ciento de significancia es -3.410 y no -1.645, este último valor sólo representa el 77,52 por ciento de la distribución.

Prueba de Raíz unitaria de Dickey-fuller aumentada

- Said y Dickey (1984) aumentaron la prueba de Dickey-Fuller básica para considerar el caso general de un ARMA(p,q) para la primera diferencia de la serie. El modelo básico esta dado,

$$y_t = \beta' D_t + \alpha y_{t-1} + u_t$$

$$\phi(L)u_t = \theta(L)\varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

- El procedimiento de la prueba consiste en estimar por MCO, la siguiente regresión $y_t = \beta' D_t + \alpha y_{t-1} + \sum_{j=1}^{j=p} \psi_j \Delta y_{t-j} + \varepsilon_t$
- Para probar la hipótesis se calcula, $ADF_t = t_{\alpha=1} = \frac{\hat{\alpha}-1}{se(\hat{\alpha})}$, y se considera este valor calculado con el valor crítico correspondiente. Si el valor calculado es mayor al valor crítico se rechaza la hipótesis nula de raíz unitaria, y por tanto el proceso sería estacionario.

Prueba de Raíz unitaria de Dickey-fuller aumentada

- Una representación alternativa de la prueba ADF, es la siguiente,

$$\Delta y_t = \beta' D_t + \rho y_{t-1} + \sum_{j=1}^{j=p} \psi_j \Delta y_{t-j} + \varepsilon_t$$

- Donde, $\rho = \alpha - 1$
- Para probar la hipótesis se calcula, $ADF_t = t_{\rho=0} = \frac{\hat{\rho}}{se(\rho)}$, y se considera este valor calculado con el valor crítico correspondiente. Si el valor calculado es mayor al valor crítico se rechaza la hipótesis nula de raíz unitaria, y por tanto el proceso sería estacionario.

Escogiendo el rezago apropiado para la prueba ADF

- Un aspecto práctico importante para implementar la prueba ADF es el número de rezagos necesarios para garantizar que el residuo sea ruido blanco.
- Si p es demasiado pequeño, entonces el residuo mantiene correlación serial, y por tanto la prueba sería sesgada.
- Si p es demasiado grande, entonces el poder de la prueba se reduce. Si el estadístico t del último rezago es mayor a 1.6, entonces $p = p_{\max}$, y proceder con la prueba de raíz unitaria, de lo contrario, reducir un rezago.
- Experimentos de montecarlo muestran que es preferible errar por exceso de rezagos que por falta de ellos.
- Una regla práctica para escoger el rezago, es utilizar el rezago máximo, $p_{\max} = \left\lceil 12 \left(\frac{T}{100} \right)^{\frac{1}{4}} \right\rceil$, donde, $[x]$ representa la parte entera de x

Prueba de Raíz Unitaria de Phillips y Perron

- La regresión auxiliar para esta prueba es la misma que la de ADF, $\Delta y_t = \beta' D_t + \rho y_{t-1} + u_t$, donde, $u_t \sim I(0)$
- La prueba de PP corrige el $t_{\rho=0}$ por la presencia de correlación serial y heterocedasticidad. Las pruebas modificadas son las siguientes,

$$Z_t = t_{\rho=0} \left(\frac{\hat{\sigma}^2}{\hat{\lambda}^2} \right)^{\frac{1}{2}} - \frac{1}{2} \left(\frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2} \right) \left(T \frac{SE(\rho)}{\hat{\sigma}^2} \right)$$

- Donde, $\hat{\sigma}^2$ y $\hat{\lambda}^2$ son estimadores consistentes de las varianzas de corto plazo.

$$\bullet \hat{\sigma}^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^{t=T} E(u_t^2). \text{ y } \hat{\lambda}^2 = \lim_{T \rightarrow \infty} \sum_{t=1}^{t=T} E(T^{-1} S_T^2)., \text{ donde,}$$

$$S_T = \sum_{t=1}^T u_t$$

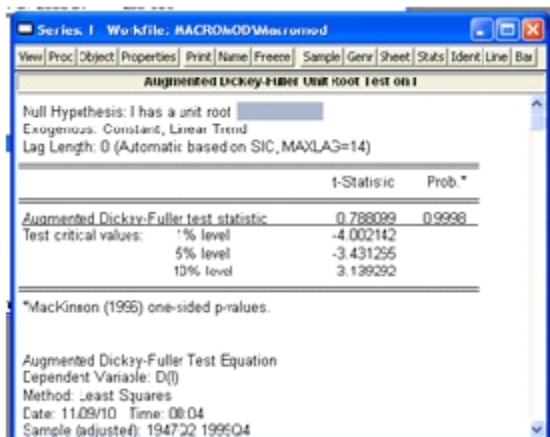
Algunos Problemas con las Pruebas de Raíz Unitaria

- Las pruebas ADF y PP son asintóticamente equivalentes pero difieren en muestras pequeñas debido a las distintas formas que utilizan para corregir la prueba por la existencia de autocorrelación serial y heterocedasticidad.
- Schwert (1989), encuentra que cuando, Δy_t es un modelo ARMA con un valor MA grande y negativo, las pruebas ADF y PP están fuertemente sesgadas, siendo la prueba PP la más distorsionada.
- ADF y PP tienen reducido poder cuando la hipótesis alternativa es un proceso autorregresivo estacionario pero que una raíz cercana a la unidad. En este caso se debe utilizar el procedimiento de selección
- El poder de las pruebas de raíz unitaria disminuye cuando existen quiebres estructurales en los componentes determinísticos o cuando no existen no-linealidades.

Ejemplo de raíz unitaria con E-views



Ejemplo de raíz unitaria con E-views



- Los modelos ARCH aparecen en los años 80 con el objetivo de modelar los clusters de volatilidad comunmente observados en series financieras.
- Los clusters de volatilidad indican que la varianza de las series analizadas no es constante en el tiempo. Por lo que es una forma de heterocedasticidad.
- También implican que los periodos de alta y baja volatilidad suelen ser persistentes.
- Estas características hacen posible que la se puedan modelar y predecir la volatilidad de las series de tiempo utilizando la propia historia de la volatilidad.

- Los momentos de los procesos ARCH han sido analizados por Engle (1982), Bollerslev (1986)
- Se dice que un proceso $\{Y_t\}$. sigue un proceso ARCH si su esperanza condicional es cero y su varianza condicional,
$$h_t^2 = g(\varepsilon_{t-1}, \varepsilon_{t-2} \dots \varepsilon_{t-k})$$
- Los modelos ARCH admiten una varianza condicional cambiante en el tiempo, pero mantienen el supuesto de varianza incondicional constante.
- El modelo ARCH(q) se define como:

$$Y_t = \varepsilon_t h_t$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i h_{t-i}^2$$

donde, $\alpha_0 > 0$, $\alpha_i \geq 0$ y $\sum_{i=1}^q \alpha_i \leq 1$

- En estos modelos la media del proceso se explica por un proceso ARMA, mientras que la varianza sigue un proceso ARCH(q)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots \alpha_q \varepsilon_{t-q}^2$$

- Una generalización del proceso anterior es el modelo GARCH, en el que,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i}^2$$

- Estos modelos permiten efectos asimétricos de los choques sobre la volatilidad de los mismos,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}^2 + \sum_{k=1}^r \gamma_k \varepsilon_{t-k}^2 I_{t-k}$$

donde, $I_t = 1$ si $\varepsilon_t < 0$, en los otros caso la variable indicador es cero.

- En este modelo choques positivos tiene efectos diferentes sobre la volatilidad que choques negativos. Choques positivos tienen un impacto, α_i , mientras que los choques negativos tienen un impacto de $\alpha_i + \gamma_i$

- El modelo EGARCH fue propuesto por Nelson (1991). En este caso el modelo es el siguiente

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$$\ln h_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \left| \frac{\varepsilon_{t-i}}{h_{t-i}} \right| + \sum_{j=1}^q \beta_j \ln h_{t-j}^2 + \sum_{k=1}^r \gamma_k \frac{\varepsilon_{t-k}}{h_{t-k}} I_{t-k}$$

- Este modelo también permite efectos asimétricos de choques siempre y cuando, $\gamma_k \neq 0$
- Todos estos modelos se estiman mediante máxima verosimilitud.

Análisis de cointegración

Definición

- Se dice que dos series, X_t y Z_t cointegran si existe una combinación lineal que es estacionaria, esto es si existe un vector $\beta = \begin{bmatrix} 1 & -\beta_2 \end{bmatrix}$, tal que $X_t - \beta_2 Z_t$ es $I(0)$
- De manera equivalente, dos series cointegran si comparten la misma tendencia estocástica, $Z_t = Z_{t-1} + \varepsilon_t$, entonces, $Z_t = Z_0 + \sum_{s=0}^{s=t} \varepsilon_s$, entonces, $X_t = \beta_2 \sum_{s=0}^{s=t} \varepsilon_s + \eta_t$
- Si X_t y Z_t cointegran, entonces existe una relación de equilibrio de largo plazo entre las series y por tanto la dinámica de corto plazo admite una representación de mecanismo de corrección de errores.

- Si X_t y Z_t no cointegran, entonces, la regresión por mínimos cuadrados ordinarios, $X_t = \beta_2 Z_t + \varepsilon_t$, es una regresión espúrea, y el residuo $\varepsilon_t \sim I(1)$
- Phillips (1986), muestra los siguientes resultados:
 - $\widehat{\beta}_2$ no converge en probabilidad al verdadero valor sino que converge a una distribución que no es normal.
 - Los estadísticos t de $\widehat{\beta}_2$ convergen a infinito.
 - El R^2 converge a 1, aún cuando no existe ninguna relación válida entre las variables.

Análisis de cointegración

Ejemplos

- Las relaciones de cointegración surgen de las condiciones de equilibrio en economía y finanzas.
- Por ejemplo, la teoría del ingreso permanente implica una relación de cointegración entre el consumo e ingreso permanente,
- El equilibrio en el mercado monetario, implica que existe una relación de cointegración entre dinero, precios, producto y la tasa de interés nominal.
- La teoría de crecimiento implica que existe una relación de cointegración entre consumo, producto e inversión, en donde la productividad es la tendencia común.
- La hipótesis de la paridad del poder de compra implica una relación de cointegración entre los precios domésticos, externos y el tipo de cambio.

Análisis de cointegración

Más ejemplos

- La ecuación de Fisher implica una relación de cointegración entre la tasa de interés nominal y la inflación.
- La teoría de expectativas de la curva de rendimiento implica cointegración entre las tasas de interés nominal a diferentes horizontes de maduración.
- El modelo de valor presente para valorizar acciones, implica que existe cointegración entre los precios de los activos y los dividendos.

Análisis de cointegración

Más ejemplos

- Las relaciones de equilibrio que la teoría económica postula se refieren a relaciones de equilibrio de largo plazo, debido a que desviaciones respecto a este equilibrio pueden tardar mucho tiempo en regresar el equilibrio. En estos casos, la cointegración se modela utilizando series medidas en la frecuencia, mensual, trimestral y anual.
- En finanzas, la cointegración puede analizarse con datos de alta frecuencia, debido a que en este caso, el concepto de cointegración esta motivado por arbitraje entre mercados, por ejemplo, la PPP:

Representación de Mecanismo de Corrección de Errores

Representación univariada

- Sean X_t y Z_t , dos variables $I(1)$ que cointegran, tal que $X_t = \beta_2 Z_t + \varepsilon_t$
- El mecanismo de corrección de errores esta dado por

$$\Delta X_t = -\gamma (X_{t-1} - \beta_2 Z_{t-1}) + \sum_{s=1}^{s=p} \Delta X_{t-s} + \eta_t$$

- Donde, $\gamma > 0$, representa la velocidad a la que se corrige el desequilibrio de corto plazo, y η_t es un ruido blanco gaussiano (con distribución normal).
- EL MCE relaciona la relación de equilibrio de largo plazo que implica la cointegración con la dinámica de corto plazo.
- Este mecanismo de corrección de errores evita que las variables de distancia mucho de su equilibrio de largo plazo.

Representación de Mecanismo de Corrección de Errores

Representación multivariada

- Sea el siguiente modelo VAR(p),

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \dots A_p X_{t-p} + \eta_t$$

- Aplicando el operador diferencias, y ordenando adecuadamente las matrices de coeficientes, se obtiene un vector de corrección de errores, VECM, de la siguiente forma,

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \dots \Gamma_{p-1} \Delta X_{t-p+1} + \eta_t$$

Donde, $\Pi = -(I - A_1 - A_2 \dots - A_p)$ y $\Gamma_i = \sum_{s=i}^{s=p-1} A_s$,

- En este modelo, si existe cointegración entonces, $\Pi = \alpha \beta'$, por tanto, $\Pi X_{t-1} = \alpha \beta' X_{t-1}$, en este caso, $\beta' X_{t-1} \sim I(0)$, y por tanto, $\Pi X_{t-1} \sim I(0)$

- Caso I: Se analiza un sólo vector de cointegración.
 - Engle y Granger (1986), la prueba se realiza en dos etapas.
- Caso II: Existen r vectores de cointegración, prueba originalmente desarrollada por Johansen (1988). El desarrolla un análisis secuencial para determinar la existencia de cointegración y el número de vectores de cointegración en base a técnicas de máxima verosimilitud.

- Se implementa en dos etapas:
 - Estimar mediante MCO, $X_t = \beta_2 Z_t + \varepsilon_t$
 - Realizar una prueba de raíz unitaria al residuo, $\hat{\varepsilon}_t = X_t - \hat{\beta}_2 Z_t$, si el residuo es estacionario entonces existe cointegración.
- La hipótesis nula es de no cointegración y la alternativa es de cointegración.

- Existen dos casos:
 - El vector de cointegración es conocido, por ejemplo porque se deriva de una relación de equilibrio, por ejemplo, la UIP. En este caso, los valores críticos de DF y PP son válidos para la prueba de raíz unitaria al residuo de la ecuación de cointegración
 - El vector de cointegración debe estimarse. En este caso la prueba de raíz unitaria al residuo de la ecuación de cointegración debe utilizar los valores críticos de Phillips y Ouliaris (1990), . C. B. Phillips and S. Ouliaris (1990): Asymptotic Properties of Residual Based Tests for Cointegration. Econometrica 58, 165–193.
- Las pruebas de cointegración utilizando un vector de cointegración pre-especificado tiene mayor poder que en el caso que el vector de cointegración se estima.

- La hipótesis a ser probada es :
$$H_0 : \varepsilon_t = \beta' Y_t \sim I(1)$$
$$H_1 : \varepsilon_t = \beta' Y_t \sim I(0)$$
- Cualquier prueba de raíz unitaria puede utilizarse para evaluar la hipótesis anterior. La elección más popular son la prueba ADF, y PP.
- Cointegración existe si la hipótesis de raíz unitaria es rechazada.
- El vector de cointegración puede incluir escalares y tendencias.

- En este caso, la hipótesis a probar es:
$$H_0 : \hat{\varepsilon}_t = \hat{\beta}'_{MCO} Y_t \sim I(1)$$
$$H_1 : \hat{\varepsilon}_t = \hat{\beta}'_{MCO} Y_t \sim I(0)$$
- En este caso, los valores críticos de la distribución ADF y PP dependen de,
 - Los términos determinísticos de la ecuación de cointegración
 - Del número de variables exógenas.

- Hansen (1992) muestra que las distribuciones de Phillips y Ouliaris (1990) dependen además de las propiedades de las variables al lado derecho e izquierdo en la ecuación de cointegración. Si la ecuación de cointegración es, $X_t = \gamma' D_t + \beta_2 Z_t + \varepsilon_t$, existen los siguientes tres casos
- Caso I: X_t y Z_t son $I(1)$ con constante, y $D_t = 1$ (Valores críticos tabla 2a, página B3 paper PO)
- Caso II Z_t tiene constante, y X_t puede o no tener constante $D_t = 1$ (Valores críticos tabla 2b, página B3 paper PO)
- Caso III Z_t tiene constante, y X_t puede o no tener constante $D_t = (1, t)$ (Valores críticos tabla 2c, página B3 paper PO)

Propiedades del vector de cointegración

- El vector de cointegración $\hat{\beta}_{MCO}$ es superconsistente, converge a su verdadero valor a la tasa T y no \sqrt{T}
- La distribución asintótica de $\hat{\beta}_{MCO}$ es sesgada asintóticamente y no normal. La fórmula del MCO para computar la varianza asintótica de $\hat{\beta}_{MCO}$ es incorrecta por lo que el t estadístico no se puede utilizar para realizar prueba de hipótesis.
- Para realizar pruebas de hipótesis cuando existe un sólo vector de cointegración se debe modificar la ecuación de largo plazo, paso 1, del método de Engle y Granger. Stock y Watson (1993) proponen un ajuste sencillo para lograr este propósito.

Mínimos Cuadrados Dinámicos: Stock y Watson (1993)

- Stock y Watson (1993) sugieren una forma simple de corregir el método de Engle y Granger para obtener un estimador eficiente del vector de cointegración.
- Sea la ecuación de cointegración, $X_t = \beta_2 Z_t + \varepsilon_t$
- Stock y Watson proponen la siguiente reparametrización, $X_t = \beta_2 Z_t + \sum_{s=-p}^{s=p} \Delta Z_{t-s} + u_t$. El estimador de MCO en la ecuación anterior se denomina estimador de mínimos cuadrados dinámicos, $\hat{\beta}_{2,MCD}$
- El estimador, $\hat{\beta}_{2,MCD}$, es consistente, eficiente y asintóticamente normal.

- La desviación estándar de $\hat{\beta}_{2,MCD}$ esta dada por la desviación estándar de MCO multiplicada por el ratio

$$f = \left(\frac{\hat{\sigma}_u^2}{\widehat{vlp}(u_t)} \right)^{\frac{1}{2}}$$

- Donde, $\hat{\sigma}_u^2 = \frac{1}{T-p} \sum u_t^2$ y $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} \dots + \phi_k u_{t-k} + \omega_t$, entonces, $vlp(u_t) = \frac{\hat{\sigma}_\omega^2}{1 - \phi_1^2 - \phi_2^2 \dots - \phi_p^2}$
- Alternativa, se puede utilizar un estimador no paramétrico de $\widehat{vlp}(u_t)$, por ejemplo el estimador de Newey-West.

$$\widehat{vlp}(u_t) = \hat{c}_0 + 2 \sum_{j=1}^k \left[1 - \frac{j}{(k+1)} \right] \hat{c}_j, \quad \hat{c}_j = T^{-1} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}$$

Vector de Corrección de Errores

- Sea Z_t un vector de k variables $I(1)$ que siguen un proceso VAR(p), la representación de corrección de errores multivariada es

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \dots \Gamma_{p-1} \Delta X_{t-p+1} + \eta_t$$

- Las propiedades de ΔX_t dependen de las propiedades de Π . Si $\text{rango}(\Pi) = 0$, entonces, $\Pi = 0$. En este caso, la dinámica de corto plazo no está relacionada con el nivel de las series, por lo que no existe cointegración, existen k tendencias estocásticas, y 0 vectores de cointegración.
- En este caso el mejor modelo para ΔX_t , está dado por,

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \dots \Gamma_{p-1} \Delta X_{t-p+1} + \eta_t$$

- Si el $\text{rango}(\Pi) = k$, entonces, $\Pi X_{t-1} \sim I(0)$, todas las series son estacionarias
- Si el $\text{rango}(\Pi) = r < k$, entonces existen r vectores de cointegración. En este caso, $\Pi = \alpha\beta'$, por lo que $\Pi X_{t-1} = \alpha_{[k \times r]} \beta'_{[r \times k]} X_{t-1}$, por lo que $\beta' X_{t-1} \sim I(0)$, con lo que $\Pi X_{t-1} \sim I(0)$.
- En este caso, existen $k - r$, tendencias estocásticas.

$AR(p)$

- Cuando existe cointegración y $\Pi = \alpha_{[k \times r]} \beta'_{[r \times k]}$, Las filas de β' forman las bases de los r vectores de cointegración, y α reparte el impacto de los vectores de cointegración en la evolución de ΔX_t , por lo que el mecanismo de corrección de errores se transforma en,

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \dots \Gamma_{p-1} \Delta X_{t-p+1} + \eta_t$$

- El vector de cointegración no es único,
 $\Pi = \alpha \beta' = \alpha H H^{-1} \beta' = (\alpha H) (\beta H^{-1})' = \alpha^* \beta'^*$
- Existen infinitos vectores de cointegración, los que se obtienen a partir de combinaciones lineales de las bases del vector de cointegración, β
- Para obtener valores únicos de α y β se requiere restricciones adicionales.

Ejemplo: Modelo VAR(1)

- Consideremos el siguiente modelo VAR(1) de variables $I(1)$,
 $Y_t = \begin{bmatrix} x_t & z_t \end{bmatrix}$, tal que

$$Y_t = A_1 Y_{t-1} + \varepsilon_t$$

- El VECM de este modelo esta dado por,

$$\Delta Y_t = \Pi Y_{t-1} + \varepsilon_t$$

$$\Pi = -(I - A_1)$$

- En este caso, si existe cointegración, $\beta'_{[2 \times 1]} = \begin{bmatrix} 1 & -\beta_2 \end{bmatrix}$, de donde,
 $\beta' Y_t = x_t - \beta_2 z_t$
- En este caso, $\text{rango}(\Pi) = 1$, y por tanto,

$$\Pi = \alpha \beta' = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -\beta_2 \end{bmatrix}$$

Ejemplo: Modelo VAR(1)

- En este caso,

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \varepsilon_t$$

- Por tanto,

$$\Delta x_t = \alpha_1 (x_t - \beta_2 z_t) + \varepsilon_{1,t}$$

$$\Delta x_t = \alpha_1 (x_t - \beta_2 z_t) + \varepsilon_{1,t}$$

- El modelo es estable, puesto que $(x_t - \beta_2 z_t)$ es estacionario.

- Los pasos básicos para aplicar esta metodología son:
 - Especificar y estimar un modelo VAR(p) para Y_t
 - Construir el ratio de verosimilitud para el $\text{rango}(\Pi)$ para determinar el número de vectores de cointegración.
 - De ser necesario, imponer restricciones de normalización e identificación a los vectores de cointegración
- Una vez normalizado e identificado los vectores de cointegración se estima el VECM por máxima verosimilitud

- Se estiman las siguientes regresiones auxiliares

$$\Delta X_t = \psi_0 + \Psi_1 \Delta X_{t-1} + \Psi_2 \Delta X_{t-2} + \dots \Psi_{p-1} \Delta X_{t-p+1} + u_t$$

$$X_{t-1} = \theta_0 + \Theta_1 \Delta X_{t-1} + \Theta_2 \Delta X_{t-2} + \dots \Theta_{p-1} \Delta X_{t-p+1} + v_t$$

- Calcular las siguientes matrices de varianza-covarianza'

$$\begin{aligned} \hat{\Sigma}_{UU} &= T^{-1} \sum_{t=1}^{t=T} u_t u_t' & \hat{\Sigma}_{UV} &= T^{-1} \sum_{t=1}^{t=T} u_t v_t' \\ \hat{\Sigma}_{VU} &= T^{-1} \sum_{t=1}^{t=T} v_t u_t' & \hat{\Sigma}_{VV} &= T^{-1} \sum_{t=1}^{t=T} v_t v_t' \end{aligned}$$

- Formar la matriz, $\hat{\Pi} = \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{UV} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{VU}$
- Encontrar los valores propios de $\hat{\Pi}$, $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots \hat{\lambda}_k$
- La función de verosimilitud del rango de $\hat{\Pi}$, esta dada por:

$$\begin{aligned} L = & - \left(\frac{Tk}{2} \right) \log(2\pi) - \left(\frac{Tk}{2} \right) - \left(\frac{T}{2} \right) \log \left| \hat{\Sigma}_{UU} \right| \\ & - \left(\frac{T}{2} \right) \sum_{i=1}^{i=h} \log(1 - \hat{\lambda}_i) \end{aligned}$$

- Las bases del vector de cointegración se determinan a partir de los r valores propios más grandes de $\hat{\Pi}$

Prueba de verosimilitud para el número de vectores de cointegración.

- Esta es una prueba secuencial,
- Se inicia la prueba, $H_0 : r = 0$ versus $H_a : r > 0$, luego, $H_0 : r = 1$ versus $H_a : r > 1$ hasta que la hipótesis nula se rechace.
- El rango de $\hat{\Pi}$ determina el número de vectores de cointegración,
- Existen dos tipos de prueba, la prueba de la traza y la del máximo valor propio.

Prueba de la traza

- En este caso, la hipótesis a testear es, $H_0 : r = r_0$ versus $H_a : r > r_0$, el estadístico de la traza esta determinado por

$$LR_{traza}(r_0) = -T \sum_{i=r_0+1}^{i=k} \ln(1 - \hat{\lambda}_i)$$

- Si $\text{rango}(\Pi) = r_0$ entonces, los valores propios, $\hat{\lambda}_{r_0+1} \dots \hat{\lambda}_k$ deberían ser todos cercanos a cero y por tanto, $LR_{traza}(r_0)$ también debería ser cercano a cero.
- Por el contrario, si $\text{rango}(\Pi) > r_0$, algunos de los valores propios, $\hat{\lambda}_{r_0+1} \dots \hat{\lambda}_k$ será diferente de cero, por lo que $LR_{traza}(r_0)$ debería ser grande.
- La distribución asintótica de $LR_{traza}(r_0)$ no es un chi-cuadrado como es estándar en la prueba del ratio de verosimilitud, sino es una versión multivariada de la distribución de Dickey-Fuller para raíz unitaria, que depende de la dimensión $k - r_0$ y de los componentes determinísticos. Los valores críticos para esta distribución han sido tabulados por Osterwald-Lenum (1992)

Prueba de la traza: Procedimiento secuencial

- Primero testear, $H_0 : r = 0$ versus $H_a : r > 0$, si la hipótesis nula no se rechaza entonces se concluye que existe cointegración.
- Si se rechaza, $H_0 : r = 0$, entonces se concluye que existe por lo menos un vector de cointegración y se procede a testear, $H_0 : r = 1$ versus $H_a : r > 1$. Si la hipótesis nula no se rechaza entonces se concluye que existe un vector de cointegración.
- Si $H_0 : r = 1$ es rechazada entonces se concluye que existe por lo menos 2 vectores de cointegración.
- El procedimiento secuencial continúa hasta que la hipótesis nula no es rechazada.

- En este caso, la hipótesis a testear es la siguiente, $H_0 : r = r_0$ versus $H_a : r = r_0 + 1$.
- En este caso, el estadístico es, $LR_{\max}(r_0) = -T \ln(1 - \hat{\lambda}_i)$
- Como en el caso del estadístico de la traza, la distribución de $LR_{\max}(r_0)$ no es chi-cuadrado sino una función complicada de procesos Browniano, que depende de $k - r_0$ y de los componentes determinísticos. Los valores críticos para esta prueba fueron tabulados por Osterwald-Lenum(1992)

- Para la prueba de cointegración de Johansen se consideran los siguientes 5 casos.
 - Sin constante, $\Delta X_t = \alpha \beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \eta_t$
 - Con constante restringida, $\Delta X_t = \alpha (\beta' X_{t-1} + \mu_0) + \Gamma_1 \Delta X_{t-1} + \eta_t$
 - Constante sin restringir $\Delta X_t = \mu_0 + \alpha (\beta' X_{t-1}) + \Gamma_1 \Delta X_{t-1} + \eta_t$
 - Tendencia restringida $\Delta X_t = \alpha (\beta' X_{t-1} + \mu_1 t) + \Gamma_1 \Delta X_{t-1} + \eta_t$
 - Constante y tendencia no restringida,
 $\Delta X_t = \mu_0 + \mu_1 t + \alpha (\beta' X_{t-1}) + \Gamma_1 \Delta X_{t-1} + \eta_t$

Propiedades del vector de cointegración

- Los vectores de cointegración no son únicos, Johansen sugiere utilizar una representación triangular, el mismo método se aplica en el Eviews.
- El vector de cointegración es asintóticamente normal y superconsistente
- Restricciones respecto al vector de cointegración se pueden testear utilizando el estadístico del ratio de verosimilitud que tiene una distribución chi-cuadrado.